



STIC Search Report

Biotech-Chem Library

File Copy
10/042059
updated

STIC Database Tracking Number: 163850

TO: David Lamberston
Art Unit: 1636
Location: rem/2B79/2C70
Serial Number: 10/042059

Thursday, October 27, 2005

From: Beverly Shears
Location: Biotech-Chem Library
REM 1A54
Phone: 571-272-2528
beverly.shears@uspto.gov

Search Notes

Protein Sequence Searches – February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension .rup) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (uniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.



STIC-Biotech/ChemLib

CRFE

168850

From: Lambertson, David
Sent: Tuesday, October 18, 2005 11:06 AM
To: STIC-Biotech/ChemLib
Cc: Lambertson, David
Subject: Search Request

Search Request

Examiner's Name:	David Lambertson
Examiner #:	79514
Art Unit:	1636
Room #:	Remsen 02B79
Mailbox room#:	Remsen 02C70
Phone:	(571) 272-0771
Results Format:	paper

Serial # 10/042059

Please Search:

Nucleic Acid databases for:

SEQ ID NO: 1 *na 843*

Including:

1. Default Search. *mej*
2. Interference Search.

Thanks,
Dave.

10/25/05
ISH.

10/18/2005

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2005, 14:13:47 ; Search time 3558 Seconds
(without alignments)
9018.600 Million cell updates/sec

Title: US-10-042-059B-1

Perfect score: 843 atggaacttctgaagccaga.....tctcctctgcttactag 843

Sequence: 1 atggaacttctgaagccaga.....tctcctctgcttactag 843

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gsa1: *
9: gb_gsa2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	299	35.5	1185	9	CNS06T5P
2	176.6	20.9	937	9	CNS06T7V
3	176.6	20.9	1052	9	CNS06XJ7
4	140	16.6	948	7	CO005472
5	137	16.3	574	8	BZ03243
6	132.8	15.8	738	8	CF877247
7	132.8	15.8	816	6	CB905545
8	125.4	14.9	788	7	CO165801
9	121.6	14.4	660	5	BX253798
10	116.8	13.9	798	6	CB619881
11	116.2	13.8	652	6	CD423256
12	113.8	13.5	747	9	CNS06T7V
13	113.2	13.4	710	9	CF475712
14	111.2	13.2	1222	3	AY103806
15	107.4	12.7	827	7	CO162312
16	106	12.6	988	1	AJ833468
17	105.8	12.6	602	7	CD982471
18	105.8	12.6	903	7	CN149106
19	104.6	12.5	527	4	CO165812
20	104.6	12.4	606	4	B1995968
21	103	12.2	737	7	CO005473
22	103	12.2	867	6	CB619882
23	102.8	12.2	727	6	CA121888
24	102.2	12.1	680	6	CA080410

25	100.8	12.0	551	5	BQ701847
26	99.4	11.8	605	7	CN913779
27	98.6	11.7	607	8	B2296833
28	98	11.6	495	1	AA786587
29	97.8	11.6	821	7	CK196770
30	97.4	11.6	837	7	CO366854
31	97	11.5	596	8	B2298767
32	97	11.5	781	7	CF667838
33	96.6	11.5	853	7	CO366784
34	96.2	11.4	657	9	AG273371
35	96	11.4	812	9	CN148636
36	95	11.3	885	7	CV274873
37	94.8	11.2	857	7	CF635400
38	94.6	11.2	480	7	CV002646
39	94.4	11.2	707	6	CA239381
40	94.2	11.2	785	7	CNS23137
41	93.6	11.1	552	4	B1718858
42	93.6	11.1	1265	3	CNS0A6WL
43	93.6	11.1	1329	3	CNS0A727
44	93.2	11.1	698	2	BE577489
45	93.2	11.1	830	6	CB622596

ALIGNMENTS

RESULT 1
LOCUS: CNS06T5P
DEFINITION: T7 end of clone AM0A029A11 of library AM0A from strain CLIB 89 of Yarrowia lipolytica, genomic survey sequence.
ACCESSION: AL414131.1 GI:12186949
VERSION: AL414131.1
KEYWORDS: GSS.
SOURCE: Yarrowia lipolytica
ORGANISM: Yarrowia lipolytica
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Dipodascaceae; Yarrowia.

REFERENCE
AUTHORS: Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Boletín-Pukhara,M., Bon,E., Broctier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durand,P., Lepingle,A., Lorente,B., Malpertuy,A., Neugebiller,C., Ozier-Kalogeropoulos,O., Portier,S., Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE: Genomic exploration of the hemiascomycetous yeasts: 1. A set of Yeast species for molecular evolution studies
JOURNAL: FEMS Lett. 487 (1), 3-12 (2000)
MEDLINE: 20584711
PUBMED: 11152876

REFERENCE
AUTHORS: Casaregola,S., Neugebiller,C., Lepingle,A., Bon,E., Feynerol,C., Artiguenave,F., Wincker,P. and Galliardin,C.
TITLE: Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia lipolytica
JOURNAL: FEMS Lett. 487 (1), 95-100 (2000)
MEDLINE: 20584727
PUBMED: 11152892

REFERENCE
AUTHORS: Direct Submission
TITLE: Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
COMMENT: This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Kluyveromyces fragilis, Kluyveromyces fragilis var. fragilis, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using BW model

Run on: October 20, 2005, 15:44:38 ; Search time 900 Seconds
(without alignments)
7726.691 Million cell updates/sec

Title: US-10-042-059B-1

Perfect score: 843
Sequence: 1 atgacattatgaagccaga.....ctccctcgtgctgactag 843

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9772231 seqs, 4124568258 residues

Total number of hits satisfying chosen parameters: 19544462

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCFUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10K_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10L_PUBCOMB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US10M_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US10N_PUBCOMB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US10O_PUBCOMB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US10P_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	843	100.0	843	US-10-042-059A-1
2	843	100.0	1655	US-10-042-059A-3
3	270	32.0	807	US-10-741-849-6032
4	156.4	18.6	869	US-10-472-587-26
5	141.2	16.7	2920	US-10-128-714-249
6	141.2	16.7	2967	US-10-128-714-5249
7	131.4	15.6	1016	US-10-437-963-93956

8	131.4	15.6	1231	9	US-09-454-279-13	Sequence 13, Appl
9	131.4	15.6	1231	17	US-10-624-061-13	Sequence 13, Appl
10	118.2	14.0	1423	18	US-10-424-599-123820	Sequence 123820,
11	116.8	13.9	658	20	US-10-653-047-6992	Sequence 6992, Ap
12	116.6	13.8	1146	20	US-10-425-114-7285	Sequence 7285, Ap
13	116	13.8	1376	20	US-10-425-115-102347	Sequence 102347,
14	112.2	13.3	1548	20	US-10-739-930-4455	Sequence 4455, Ap
15	111.2	13.2	1223	9	US-09-454-279-11	Sequence 11, Appl
16	111.2	13.2	1223	17	US-10-624-061-11	Sequence 11, Appl
17	111.2	13.2	1500	20	US-10-767-101-14553	Sequence 14553, A
18	110.6	13.1	1010	20	US-10-425-115-102351	Sequence 102351,
19	109.6	13.0	1397	20	US-10-425-115-102346	Sequence 102346,
20	104.4	12.4	1217	14	US-10-267-763-4	Sequence 4, Appl
21	104.4	12.4	1217	14	US-10-267-763-10	Sequence 10, Appl
22	103.2	12.2	1368	20	US-10-425-115-102349	Sequence 102349,
23	99.8	11.8	1020	9	US-09-454-279-15	Sequence 15, Appl
24	99.8	11.8	1020	17	US-10-624-061-15	Sequence 15, Appl
25	99.8	11.8	1079	18	US-10-425-114-7639	Sequence 7639, Ap
26	97.4	11.6	798	9	US-09-938-842A-972	Sequence 972, App
27	97.4	11.6	798	11	US-09-938-842A-972	Sequence 972, App
28	97.4	11.6	1006	14	US-10-267-763-3	Sequence 3, Appl
29	97.4	11.6	1006	14	US-10-267-763-9	Sequence 9, Appl
30	97.4	11.6	1006	14	US-10-267-763-8	Sequence 8, Appl
31	93.6	11.1	1430	20	US-10-739-930-4454	Sequence 4454, Ap
32	93.6	11.1	1005	9	US-09-938-842A-737	Sequence 737, App
33	93.6	11.1	1005	11	US-09-938-842A-737	Sequence 737, App
34	93.6	11.1	1207	14	US-10-267-763-8	Sequence 8, Appl
35	93.6	11.1	1207	14	US-10-267-763-8	Sequence 8, Appl
36	87.6	10.4	1653	18	US-10-424-599-93036	Sequence 93036, A
37	81	9.6	880	18	US-10-425-114-3898	Sequence 3898, Ap
38	80	9.5	1314	20	US-10-425-115-151470	Sequence 151470,
39	78.4	9.3	1197	18	US-10-425-114-72576	Sequence 72576, A
40	77.2	9.2	524	20	US-10-425-115-102342	Sequence 102342,
41	76.2	9.0	768	19	US-10-437-963-89372	Sequence 89372, A
42	74.8	8.9	2154	19	US-10-437-963-83557	Sequence 83557, A
43	74.2	8.8	1066	18	US-10-425-114-1332	Sequence 1332, Ap
44	74.2	8.8	1083	20	US-10-425-115-115559	Sequence 115559,
45	73	8.7	580	19	US-10-021-323-12231	Sequence 12231, A

ALIGNMENTS

RESULT 1
US-10-042-059A-1
; Sequence 1, Application US/10042059A
; Publication No. US20020197704A1
; GENERAL INFORMATION:
; APPLICANT: Rneth Biotech Gesellschaft fur neue biotechnologische Prozesse und
; APPLICANT: Protektur mbH
; TITLE OF INVENTION: Nucleic Acid Molecule Containing a Nucleic Acid Coding for a Pol.
; TITLE OF INVENTION: with Chorismate Mutase Activity
; FILE REFERENCE: PA10558US-019
; CURRENT APPLICATION NUMBER: US/10/042, 059A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: DE 199 19 124.7
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Hansenula polymorpha
US-10-042-059A-1

Query Match 100.0%; Score 843; DB 13; Length 843;
Best Local Similarity 100.0%; Pred. No. 8.7e-267;
Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACCTTATGAGCAGCAAGACAGTGTGACCTTGGCAACTTGAAGATGCTTGGTC 60
DB 1 ATGACCTTATGAGCAGCAAGACAGTGTGACCTTGGCAACTTGAAGATGCTTGGTC 60

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: October 20, 2005, 14:14:27 ; Search time 198 Seconds
(without alignments)
6966.572 Million cell updates/sec

Title: US-10-042-059B-1

Perfect score: 843
Sequence: 1 atgaccttatagaagccaga.....tctcctcgtcgtctactag 843

Scoring table:
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270	32.0	948	4	US-09-248-796A-4110 Sequence 4110, Ap
2	131.4	13.6	1231	4	US-09-454-279-13 Sequence 13, Appl
3	111.2	12.3	1223	4	US-09-454-279-11 Sequence 11, Appl
4	104.4	12.4	1217	3	US-09-610-040-4 Sequence 4, Appl
5	104.4	12.4	1217	3	US-09-610-040-10 Sequence 10, Appl
6	104.4	12.4	1217	4	US-10-267-763-4 Sequence 4, Appl
7	104.4	12.4	1217	4	US-10-267-763-10 Sequence 10, Appl
8	99.8	11.8	1020	4	US-09-454-279-15 Sequence 15, Appl
9	97.4	11.6	1006	3	US-09-610-040-3 Sequence 3, Appl
10	97.4	11.6	1006	3	US-09-610-040-9 Sequence 9, Appl
11	97.4	11.6	1006	4	US-10-267-763-3 Sequence 3, Appl
12	97.4	11.6	1006	4	US-10-267-763-9 Sequence 9, Appl
13	93.6	11.1	1207	3	US-09-610-040-2 Sequence 2, Appl
14	93.6	11.1	1207	3	US-09-610-040-8 Sequence 8, Appl
15	93.6	11.1	1207	4	US-10-267-763-2 Sequence 2, Appl
16	93.6	11.1	1207	4	US-10-267-763-8 Sequence 8, Appl
17	61	7.2	780	4	US-09-454-279-17 Sequence 17, Appl
18	60.8	7.2	7218	1	US-08-232-463-14 Sequence 14, Appl
19	59.8	7.1	579	4	US-09-454-279-7 Sequence 7, Appl
20	53.4	6.3	525	4	US-09-454-279-5 Sequence 5, Appl
21	37.6	4.5	4079	4	US-09-016-434-1248 Sequence 1248, Ap
22	37.6	4.5	4517	4	US-09-949-016-4573 Sequence 4573, Ap
23	37.6	4.5	4519	4	US-09-023-655-1202 Sequence 1202, Ap
24	37.6	4.5	4519	4	US-09-949-016-650 Sequence 650, Appl
25	37.2	4.4	1590	4	US-09-492-027-3 Sequence 3, Appl
26	37.2	4.4	15307	4	US-09-252-991A-15307 Sequence 15307, A
27	37.2	4.4	2352	4	US-09-252-991A-15332 Sequence 15332, A

28	37.2	4.4	2916	4	US-09-252-991A-15259 Sequence 15259, A
29	36.8	4.4	2169	3	US-09-434-408-3 Sequence 3, Appl
30	36.6	4.3	5176	3	US-09-610-040-6 Sequence 6, Appl
31	36.6	4.3	5176	4	US-10-267-763-6 Sequence 6, Appl
32	36.4	4.3	601	4	US-09-454-279-3 Sequence 3, Appl
33	35.4	4.2	4403765	3	US-09-103-840A-2 Sequence 2, Appl
34	35.4	4.2	4411529	3	US-09-103-840A-1 Sequence 1, Appl
35	35	4.2	2097	4	US-09-252-991A-13635 Sequence 13635, A
36	35	4.2	2331	4	US-09-252-991A-13809 Sequence 13809, A
37	34.6	4.1	789	4	US-09-252-991A-12701 Sequence 12701, A
38	34.6	4.1	1632	4	US-09-252-991A-12701 Sequence 12701, A
39	34.4	4.1	3008	4	US-09-435-376-6 Sequence 6, Appl
40	34.2	4.1	1509	4	US-09-724-797-89 Sequence 89, Appl
41	34.2	4.1	11958	3	US-09-134-246-8 Sequence 8, Appl
42	34.2	4.1	11958	4	US-09-664-186-8 Sequence 8, Appl
43	34	4.0	705	4	US-09-270-767-5061 Sequence 5061, Ap
44	34	4.0	705	4	US-09-270-767-20343 Sequence 20343, A
45	34	4.0	978	4	US-09-902-540-6442 Sequence 6442, Ap

ALIGNMENTS

RESULT 1
US-09-248-796A-4110
Sequence 4110, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 4110
LENGTH: 948
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-4110

Query Match 32.0%; Score 270; DB 4; Length 948;
Best Local Similarity 59.1%; Pred. No. 1.4e-72;
Matches 462; Conservative 0; Mismatches 320; Indels 0; Gaps 0;

Qy	1	ATGACCTTATGACGCGAAGAAACAGTCTGACCTTGGCAATTAAGATCCCTTGC	60
Db	142	ATGACCTTATGACGCGAAGAAACAGTCTGACCTTGGCAATTAAGATCCCTTGC	201
Qy	61	CGGATGAGGATGACATCATCTTCACTTATGAGCGGTGCAATCTATGCGTCC	120
Db	202	ACGATGAGGATGACATCATCTTCACTTATGAGCGGTGCAATCTATGCGTCC	261
Qy	121	TGGTATTAACAAGTCAACAGTCTTCCATTTCCCACTTGGAGCGTCTTGGACTG	180
Db	262	TGAGTTATGAAAGATTAATATATATCCCAATTTTGAAGAACTTTTGGAAATG	321
Qy	181	CTGTGTGCGACGACGAGGAAATCAATTCGACGTTGAGAGATGACCGCCAGACG	240
Db	322	GCTTGTATTAAGTGAAGTGTCTATTTCAATTCAGCGTTATGAGGACACAGCGA	381
Qy	241	GTCCTTTTTCGCAAGTGTGCAAGGTTTCCCAAGTCACTACCATCG	300
Db	382	ACTCAATTTTTCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	441
Qy	301	GTCGACCTCTTACGAGGATGAAATCAACGTTCAACAGATTAAGTTCACAG	360
Db	442	ATATTGGCAATATTCGATGAATTAATGTTAATTCGAAATATGAAATGTTATGTT	501

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2005, 14:09:57 ; Search time 591 Seconds
(without alignments)
8443.895 Million cell updates/sec

Title: US-10-042-059B-1

Perfect score: 843
Sequence: 1 atggacttatgaagccaga.....tctcctcgtgcttactag 843

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_16Dec04:*
2: geneeqn1980s:*
3: geneeqn1990s:*
4: geneeqn2000s:*
5: geneeqn2001as:*
6: geneeqn2002as:*
7: geneeqn2003as:*
8: geneeqn2004as:*
9: geneeqn2005as:*
10: geneeqn2006as:*
11: geneeqn2007as:*
12: geneeqn2008as:*
13: geneeqn2009as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	843	100.0	843	3	AAC81949
2	843	100.0	1655	3	AAC81950
3	270	32.0	807	12	ADP98547
4	182.8	21.7	801	13	ADR85510
5	156.4	16.7	869	8	AAL50202
6	141.2	16.7	989	13	ADR84923
7	141.2	16.7	2920	8	ABT17891
8	141.2	16.7	2987	8	ABT19705
9	141.2	16.7	6989	13	ADR84336
10	131.4	15.6	1231	10	ABX93057
11	131.4	15.6	1231	12	ADJ55263
12	116.8	13.9	658	3	AAR14469
13	111.2	13.2	1223	10	ABX93056
14	111.2	13.2	1223	12	ADJ55261
15	104.4	12.4	1217	6	ABA91395
16	104.4	12.4	1217	6	ABA91401
17	99.8	11.8	1020	10	ABX93058
18	99.8	11.8	1020	12	ADJ55265
19	97.4	11.6	798	6	ABZ13167
20	97.4	11.6	993	6	ABA99614

21	97.4	11.6	1006	6	ABA99613	ABA99613 A. thalia
22	97.4	11.6	1006	6	ABA91400	ABA91400 Arabidops
23	97.4	11.6	1006	6	ABA91394	ABA91394 Arabidops
24	93.6	11.1	1005	6	ABZ12932	ABZ12932 Arabidops
25	93.6	11.1	1207	6	ABA91393	ABA91393 Arabidops
26	93.6	11.1	1207	6	ABA91399	ABA91399 Arabidops
27	93.6	11.1	1274	3	AAC33324	AAC33324 Arabidops
28	73	8.7	580	13	ACN57510	ACN57510 Cotton gy
29	69.8	8.3	622	13	ACN50132	ACN50132 Cotton no
30	64.2	7.6	600	13	ADR64815	ADR64815 Cotton CD
31	61.8	7.3	258	6	ABT17526	ABT17526 Corn taas
32	61	7.2	780	12	ABX93059	ABX93059 Wheat cho
33	61	7.2	780	12	ADJ55267	ADJ55267 Wheat cho
34	59.8	7.1	579	10	ABX93054	ABX93054 Wheat cho
35	59.8	7.1	579	12	ADJ55257	ADJ55257 Wheat cho
36	56.2	6.7	584	13	ACN50205	ACN50205 Cotton no
37	54.8	6.5	542	13	ACN57601	ACN57601 Cotton gy
38	53.4	6.3	525	10	ABX93053	ABX93053 Soybean c
39	53.4	6.3	525	12	ADJ55255	ADJ55255 Soybean c
40	52.6	6.2	524	13	ADR64235	ADR64235 Cotton CD
41	49.4	5.9	2000	8	ADR71938	ADR71938 Rice gene
42	48.8	5.8	864	12	ADJ44958	ADJ44958 Plant CDN
43	47.6	5.6	1379	6	ABA91794	ABA91794 Yeast mlt
44	47	5.6	5858	8	ADJ55726	ADJ55726 Nephtila m
45	46.8	5.6	180385	10	ADL13931	ADL13931 Osteoarth

ALIGNMENTS

RESULT 1	
ID	AAC81949 standard; DNA; 843 BP.
XX	
AC	AAC81949;
XX	
DT	15-SEP-2003 (revised)
DT	01-MAR-2001 (first entry)
XX	
DE	H. polymorpha chorismate mutase DNA.
XX	
KW	Chorismate mutase; prephenate; selection marker; auxotrophic yeast; ds.
XX	
OS	Pichia angusta.
XX	
PN	WO200065071-A1.
XX	
PD	02-NOV-2000.
XX	
PF	27-APR-2000; 2000WO-EP003844.
XX	
PR	27-APR-1999; 99DE-01019124.
XX	
PA	(RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX	
PI	Gellissen G, Braus G, Pries R, Krappmann S, Strasser AW,
XX	
XX	F-PDSB; AAB11451.
DR	WPI; 2000-687355/67.
XX	
PT	Nucleic acids encoding chorismate mutase, useful for preparing an
PT	auxotrophic selection system for recombinant yeast and for recombinant
XX	protein expression.
XX	
PS	Claim 1a; Page 57; 63pp; German.
XX	
CC	This invention describes novel nucleic acids (I) that encode a
CC	polypeptide (II) with chorismate mutase (CM) activity (or its
CC	complementary strand). CM catalyzes conversion of chorismate to
CC	prephenate, an essential precursor for Phe and Tyr. (I) is a selection
CC	marker for construction of corresponding auxotrophic yeast (requiring Phe
CC	and Tyr) that are used for recombinant production of proteins. (I) allows
CC	selection of transformed yeast on simple media. (Updated on 15-SEP-2003

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 20, 2005, 14:12:47 : Search time 4117 Seconds
(without alignments)
9921.726 Million cell updates/sec

Title: US-10-042-059B-1

Perfect score: 843
Sequence: 1 atgagcttataagagcagca.....tctctctgcgtctgactag 843

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBml: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	843	100.0	843	BD266599 Nucleic a
2	843	100.0	843	BD266599 Nucleic a
3	843	100.0	1652	AF204738 Pichia an
4	843	100.0	1655	BD266600 Nucleic a
5	843	100.0	1655	AX043760 Sequence
6	325.2	38.6	110000	Continuation (21 o
7	281.8	32.4	110000	Continuation (21 o
8	270	32.0	948	AR548979 Sequence
9	256.4	30.4	110000	Continuation (2 of
10	217.4	25.8	2059	YSCAR07A
11	217.4	25.8	43776	SCCHXVI
12	217.4	25.8	16536	AY693179
13	217.4	25.8	110000	Continuation (8 of
14	210.4	22.4	110000	Continuation (12 o
15	189	18.6	869	BD178335
16	156.4	16.8	1348	AB116236
17	141.8	15.6	1231	AR404687
18	141.8	15.6	1231	AR404687
19	131.4	15.6	1231	AR404687

20	131.4	15.6	1250	8	AK068983	ORYZA SAT
21	113.6	13.5	3167	8	AF133241	EMERICELL
22	111.2	13.2	1223	6	AR404686	SEQUENCE
23	110	13.0	1000	8	L47356	LYCOPERISCO
24	107.6	12.8	951	8	BT005306	ARABIDOPSIS
25	107.6	12.8	1143	8	AK117860	ARABIDOPSIS
26	104.4	12.4	1217	6	AR236635	SEQUENCE
27	104.4	12.4	1217	6	AX36641	SEQUENCE
28	104.4	12.4	1217	6	AX464576	SEQUENCE
29	104.4	12.4	1217	6	AX464582	SEQUENCE
30	104.4	12.4	1217	6	AF131219	ARABIDOPSIS
31	100.6	11.9	110000	2	AP006499	OF
32	99.8	11.8	1020	6	AR404688	SEQUENCE
33	97.4	11.6	798	6	AX506277	SEQUENCE
34	97.4	11.6	829	8	AY133840	ARABIDOPSIS
35	97.4	11.6	993	6	AX343935	SEQUENCE
36	97.4	11.6	1006	6	AR236634	SEQUENCE
37	97.4	11.6	1006	6	AR236640	SEQUENCE
38	97.4	11.6	1006	6	AX464575	SEQUENCE
39	97.4	11.6	1006	6	AX464581	SEQUENCE
40	97.4	11.6	1006	6	L47355	ARABIDOPSIS
41	97.4	11.6	1006	6	AX464581	SEQUENCE
42	97.4	11.6	1055	8	AY065238	SEQUENCE
43	93.6	11.1	1005	6	AX506042	SEQUENCE
44	93.6	11.1	1207	6	AR236633	SEQUENCE
45	93.6	11.1	1207	6	AR236639	SEQUENCE

ALIGNMENTS

RESULT 1
LOCUS BD266599
DEFINITION Nucleic acid molecule, containing a nucleic acid which codes for a polypeptide with chorismate mutase activity.

ACCESSION BD266599
VERSION BD266599.1 GI:33076367
KEYWORDS JP 2002542788-A/1.
SOURCE Pichia angusta
ORGANISM Pichia angusta

REFERENCE 1 (bases 1 to 843)
AUTHORS Gellissen, G., Braus, G., Pries, R., Krappmann, S. and Strasser, A.W.
TITLE Nucleic acid molecule, containing a nucleic acid which codes for a polypeptide with chorismate mutase activity
JOURNAL Patent: JP 2002542788-A 1 17-DEC-2002;
RHEIN BIOTECH GESELLSCHAFT FUER NEUE BIOTECHNOLOGISCHE PROZESSE UND

COMMENT OS Hansenula polymorpha (yeast)
PN JP 2002542788-A/1
PD 17-DEC-2002
PF 27-APR-2000 JP 2000614405
PI 27-APR-1999 DE 199 19 124.7
PR GERR GELLISSEN, GERHARD BRAUS, RALPH PRIES, SVEN KRAPPMANN, PI

PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10 PC
PC C12N9/90, C12N15/00,
PC C12N5/00
CC Nucleic acid molecule, containing a nucleic acid which codes
CC for a
CC polypeptide with chorismate mutase activity
CC Key
FT source
FT 1. 843
FT location/Qualifiers
FT 1. 843
FT /organism="Hansenula polymorpha (yeast)"

FEATURES
source
1. 843
/organism="Pichia angusta"
/mol_type="genomic DNA"
/db_xref="taxon:4905"

ORIGIN